

HERRAMIENTA PARA ANÁLISIS DE LA LEGIBILIDAD LINGÜÍSTICA DE CONTENIDO WEB EN ESPAÑOL COMO APOYO A PROCESOS DE EVALUACIÓN DE ACCESIBILIDAD

TOOL FOR ANALYSIS OF THE LINGUISTIC LEGIBILITY OF WEB CONTENT IN SPANISH AS SUPPORT TO ACCESSIBILITY ASSESSMENT PROCESSES

Diana Patricia Quintero Lorza

Administradora de Sistemas Informáticos de la Universidad Nacional de Colombia (Manizales/Colombia).
E-mail: dpquinterol@unal.edu.co

Valentina Tabares Morales

Doctora en Ingeniería de la Universidad Nacional de Colombia (Manizales/Colombia). Profesora Asociada de la Universidad Nacional de Colombia (Manizales/Colombia).
E-mail: vtabaresm@unal.edu.co

Néstor Darío Duque Méndez

Doctor en Ingeniería de la Universidad Nacional de Colombia (Manizales/Colombia). Profesor Titular de la Universidad Nacional de Colombia (Manizales/Colombia).
E-mail: ndduqueme@unal.edu.co

Recebido em: 7 de abril de 2021
Aprovado em: 16 de junho de 2021
Sistema de Avaliação: Double Blind Review
BCIJ | v. 1 | n. 1 | p. 71-87 | jul./dez. 2021
DOI: <https://doi.org/10.25112/bcij.v1i1.2677>



RESUMEN

El ejercicio de la comprensión lectora es una práctica cotidiana e imperativa de la configuración del mundo contemporáneo como parte de los aprendizajes autónomos, lo cual ha generado que la sociedad actual se caracterice por interpretar o producir textos y contenidos generalmente ubicados en sitios web. La legibilidad lingüística se puede considerar como una condición necesaria para que un texto se pueda leer y se facilite su comprensión. Por otro lado, la accesibilidad web es un concepto que hace referencia al conjunto de elementos que permiten a un usuario sin importar sus condiciones pueda acceder a los contenidos ofrecidos en la web. Cuando se realizan procesos de evaluación de accesibilidad es importante determinar el nivel de complejidad del contenido textual, para determinar si es necesario realizar cambios, entregar otras alternativas del texto o material complementario. Sin embargo, determinar esto no es un proceso sencillo y puede requerir del análisis del nivel de legibilidad de los textos. En este trabajo se propone una herramienta denominada Hohkma, que permite analizar contenido textual en español, para determinar su nivel de legibilidad utilizando los índices de Fernández Huerta y Flesch-Szigriszt. La herramienta entrega explicación de los resultados, estadísticas y recomendaciones, con un enfoque desde el punto de vista de la accesibilidad, con el objetivo de convertirse en una alternativa que apoye en procesos de evaluación orientados a lograr sitios web más accesibles.

Palabras clave: Legibilidad. Accesibilidad. Contenido Web.

ABSTRACT

The exercise of reading comprehension is a daily and imperative practice of the configuration of the contemporary world as part of autonomous learning, which has generated that today's society is characterized by interpreting or producing texts and content generally located on websites. Linguistic readability can be considered as a necessary condition for a text to be readable and easy to understand. On the other hand, web accessibility is a concept that refers to the set of elements that allow a user, regardless of its conditions, to access the content offered on the web. When accessibility evaluation processes are carried out, it is important to determine the level of complexity of the textual content, to determine if it is necessary to make changes, deliver other alternatives to the text or complementary material. However, determining this is not an easy process and may require analysis of the level of readability of the texts. In this work, a tool called Hohkma is proposed, which allows analyzing textual content in Spanish, to determine its level of readability using the Fernández Huerta and Flesch-Szigriszt indices. The tool provides an explanation of the results, statistics and recommendations, with a focus from the point of view of accessibility, with the aim of becoming an alternative that supports evaluation processes aimed at achieving more accessible websites.

Keywords: Readability. Accessibility. Web content..



1 INTRODUCCIÓN

El ejercicio de la comprensión lectora es una práctica cotidiana e imperativa de la configuración del mundo contemporáneo como parte de los aprendizajes autónomos, lo cual ha generado que la sociedad actual se caracterice por interpretar o producir textos y contenidos generalmente ubicados en sitios web. De este modo, los aprendizajes autónomos a lo largo de la vida de un individuo se fundamentan en la comprensión lectora (BAGUMA; LUBEGA, 2008) business, personal efficiency or even improvement of livelihoods [1], Web users including People with Disabilities (PWDs. Bajo este enfoque, cada vez es más necesario que la elaboración del contenido se adapte y responda a un saber disciplinario antes que uno comercial. Hay exigencias requeridas desde diversos enfoques (GARCÍA; ARÉVALO; HERNÁNDEZ SUÁREZ, 2018): la teoría del diseño curricular la cual espera la alineación del texto de acuerdo con un tema específico, la teoría del diseño tipográfico como parte de la estructura grafica del texto y la teoría del diseño lingüístico el cual exige la incorporación de las propiedades textuales y de legibilidad.

La legibilidad de los textos escritos es un tema de principal interés para las personas encargadas de generar, enseñar, promover y evaluar contenido textual con el objetivo de afianzar la capacidad de lectura o utilizan la escritura como medio de difusión. Generalmente estas personas deben tomar decisiones acerca del material que van a generar, utilizar o difundir, pero estas decisiones pueden llegar a ser erróneas a falta de elementos de juicio (ALLIENDE, 1990). Estas situaciones han llevado a la búsqueda de procedimientos objetivos que permitan realizar una medición de la legibilidad de los textos escritos y proporcionar a los interesados una información confiable. En el siglo pasado la elaboración de fórmulas de legibilidad tuvo gran desarrollo a partir de 1950, especialmente para la lengua inglesa y aunque en menor medida para el español, el francés y otros idiomas (GUAMÁN CASTILLO, 2014). En la actualidad se descarta el uso de las fórmulas como único modo de medir la legibilidad lingüística de un texto; sino como un procedimiento entre otros que permiten dar una base para formular un juicio. La necesidad de contar con instrumentos que ayuden a determinar adecuadamente la legibilidad de un texto para determinados grupos de lectores es particularmente urgente en procesos de evaluación de accesibilidad web.

Por otro lado, la accesibilidad web es un concepto que hace referencia al conjunto de elementos que permiten a un usuario, sin importar sus condiciones, que pueda acceder a los contenidos ofrecidos en la web (MIÑÓN *et al.*, 2014) accessibility is nevertheless quite difficult to achieve, let alone to maintain at a high level of quality throughout the development process. This difficulty is due to diverse factors including, but not limited to, the lack of proper development methods, authoring tools and accessibility training for user interface development professionals. In an attempt to offer a solution to these difficulties, this paper proposes a methodological approach for the integration of accessibility requirements into a user interface



development method based on User Interface Description Language (UIDL). Un sitio web es considerado accesible cuando desde su diseño se han considerado las características físicas y cognitivas de los usuarios (ANTONELLI; FORTES, 2015).

Sin embargo, el análisis de legibilidad lingüística no es usualmente evaluado en los procesos de análisis para determinar si un sitio web es accesible o no de acuerdo con el nivel educativo o el público al que está enfocado el contenido textual del mismo. Uno de los factores por los que se genera esta problemática, es debido al hecho de que los desarrolladores desconocen los lineamientos para el desarrollo de material lingüísticamente accesible o en el caso de que se conozcan, se dificulta su aplicación en términos prácticos debido a la complejidad con que se expresan.

Teniendo en cuenta lo anterior, en este trabajo se desarrolla una herramienta web para la evaluación de legibilidad lingüística en español como apoyo durante los procesos de evaluación de accesibilidad de sitios web, esta herramienta es creada en base a las pautas de accesibilidad web definidas en las normas e iniciativas internacionales, una revisión de los métodos y herramientas tecnológicas actualmente utilizados en el análisis de legibilidad y la identificación de las necesidades que puedan llegar a presentar los evaluadores durante este proceso.

En la siguiente sección de este artículo se presentan conceptos clave asociados con la problemática abordada. En la sección 2 se presentan algunos trabajos relacionados, seguida del planteamiento del problema y la propuesta. Finalmente se presentan las conclusiones y trabajo futuro.

2 CONCEPTOS

La **legibilidad** es definida como “un atributo del texto, un conjunto de peculiaridades que favorecen o dificultan la comprensión lectora” (BARRIO-CANTALEJO *et al.*, 2008). Se puede considerar como una condición necesaria para que un texto se pueda leer y se facilite su comprensión, siendo una característica que puede favorecer o dificultar la comunicación efectiva entre los escritores y los lectores (HERNÁNDEZ, 2010)(ALLIENDE, 1990).

Es posible entender la legibilidad desde dos corrientes, la primera es la legibilidad material o tipográfica y la segunda corresponde a la legibilidad lingüística, que hace referencia a las condiciones léxicas y gramaticales de un texto. En idiomas como el inglés la legibilidad se estudia únicamente bajo el primer aspecto, sin embargo, para el español el término legibilidad comprende la integración de ambos. Para este trabajo, se tiene en cuenta la legibilidad lingüística, que permite analizar la dificultad del texto (vocabulario conocido o desconocido) y los aspectos morfosintácticos, como las construcciones gramaticales utilizadas que permiten que un lector enfrente exitosamente el texto (ALLIENDE, 1990).



Las Fórmulas de Legibilidad, comprenden uno de los métodos para medir la legibilidad lingüística. Facilitan la recepción del mensaje verbal a partir de condiciones de producción textual, entre las fórmulas para la evaluación del idioma español se encuentran las siguientes:

- **Índice de Fernández Huerta:** también llamada fórmula de "Lecturabilidad", planteada por José Fernández Huerta en 1959 (RIOS HERNANDEZ, 2017) y es considerado uno de los pioneros en hablar de la legibilidad lingüística para el castellano. Este índice es comúnmente utilizado debido a su aplicabilidad para todo tipo de texto que esté compuesto de mínimo 100 palabras. Para este índice se valora el texto en una escala de 100 puntos; los textos con un índice de legibilidad de 51 puntos o menos, es considerado entre un poco difícil a muy difícil, asociados a niveles académicos a partir del bachillerato (10° y 11° grado); mientras que se aceptan como normal los textos con valores de 61 a 70 puntos y los textos con evaluaciones de 71 o más están asociados con grados de escolaridad de 6° o menos (FERNÁNDEZ; MANRIQUE-ABRIL; BAUTISTA SAAVEDRA, 2010).
- **Legibilidad de μ :** es propuesta por Miguel Muñoz Baquedano y José Muñoz Urra en el año 2006. Los cuales establecen que la legibilidad de un texto está determinada por "la varianza total de caracteres de un texto en función de las distancias al cuadrado entre cada carácter y el promedio de caracteres de la frecuencia" (MUÑOZ BAQUEDANO, 2006). Es decir, a una menor variabilidad en el total de caracteres, el grado de dificultad del texto disminuye y a mayor variabilidad de caracteres, mayor la dificultad de lectura del texto. Este modelo de legibilidad se basa en que la "riqueza del vocabulario" está relacionado con la "dispersión de los puntajes respecto al promedio de caracteres de las palabras en un texto" (MUÑOZ BAQUEDANO, 2006). De igual forma que con otros modelos de legibilidad, los autores proponen una escala de grados de legibilidad con el propósito asociar los resultados del modelo a un nivel de dificultad de lectura en una escala de grados de 0 a 100.
- **Índice de Flesch-Szigriszt:** conocido como la Fórmula de Perspicuidad fue creada por Francisco Szigriszt Pazos en 1992 para su tesis doctoral, en la cual diseñó dos fórmulas para análisis de legibilidad, la primera para aquellos textos con menos de 100 palabras y la segunda versión, la cual será utilizada para este trabajo, es válida para textos de 100 palabras o más (BALLESTEROS-PEÑA; FERNÁNDEZ-AEDO, 2013). Esta es una de las fórmulas más utilizadas al momento de realizar análisis de legibilidad debido a que es aplicable sobre textos de diversos enfoques temáticos. El índice indica que un texto presenta una legibilidad normal cuando obtiene una valoración entre 51 y 65, su complejidad aumenta si su valoración desciende, en los intervalos más bajos ([0-15]) se clasifican aquellos textos considerados de índole científica y filosófica.



Mientras si la valoración aumenta la complejidad del texto disminuye, se tiene en cuenta que el índice presenta una escala de 0 a 100 (PORRAS-GARZÓN; ESTOPÀ, 2020).

- **Escala de Inflesz:** En (BARRIO-CANTALEJO et al., 2008) se establece que en la Fórmula de Perspicuidad existe una falta de consistencia al establecer los grados de dificultad, ya que considera que esta no es representativa, por lo que se implementa como método de interpretación a los resultados la escala de Inflesz, la cual establece cinco niveles de dificultad, asignando a cada nivel una tipología textual.
- **Fórmula de Crawford:** esta fórmula es propuesta por Alan N. Crawford en 1989, con el objetivo de calcular los años de escolaridad necesarios para entender un texto, pero solo es válida para los grados de escolaridad que comprenden la primaria. Es útil para aquellos que escriben materiales didácticos en español para primaria y para aquellos que desean clasificar los materiales ya existentes.

Por otro lado, la **accesibilidad web** procura que los contenidos y servicios disponibles a través de la Web puedan ser utilizados por todas las personas, por lo que existe la necesidad de vencer las barreras impuestas al momento de crear estos elementos (ARAGALL, 2010). Se puede entender la accesibilidad web como la facilidad con que un usuario puede acceder al contenido de un sitio web independientemente de sus conocimientos y condiciones personales (HILERA *et al.*, 2013).

Con el fin de lograr que se cuente con sitios web accesibles, el World Wide Web Consortium (W3C) propuso un conjunto de Pautas de Accesibilidad al Contenido Web (WCAG por sus iniciales en inglés), ofreciendo una serie de recomendaciones que permiten evaluar contenidos web bajo los principios: perceptible, operable, comprensible y robusto para determinar su nivel de accesibilidad (KIRKPATRICK ANDREW, 2018)(FIGUEROA *et al.*, 2014). Además, entrega a los desarrolladores lineamientos para que al momento de crear los sitios web, puedan entender a los distintos usuarios en su diversidad (SALAZAR GRIJALVA, 2019).

Dentro de las WCAG se encuentra la pauta de accesibilidad 3.1 (legible) y dentro de esta se encuentra el criterio 3.1.5 en el cual se hace referencia al nivel de educación requerida para la comprensibilidad de un texto. Debido al hecho de que cada usuario experimenta un proceso de lectura de manera diferente, existe la posibilidad de que un usuario encuentre dificultades al momento de reconocer una palabra escrita o comprender un documento. La intención de esta guía es garantizar a los usuarios que la información necesaria para comprender un texto se encuentre disponible (W3C, 2010).

Teniendo en cuenta lo anterior, en un proceso de evaluación de accesibilidad es de importancia poder determinar el nivel de legibilidad de un texto, con el fin de dar cumplimiento al criterio 3.1.5 de las WCAG,



el cuál indica que cuando un texto requiera un nivel de lectura más avanzado, debe existir algún contenido complementario o una versión que no requiera una capacidad de lectura más avanzada. (W3C, 2008)

3 TRABAJOS RELACIONADOS

Se realizó una exploración de trabajos relacionados, con el enfoque de la utilización de los modelos cuantitativos de evaluación de legibilidad de los textos y el contenido web, mediante la utilización de herramientas web, offline o analógico. Empleado en diversos temas de interés con el propósito de identificar las metodologías propuestas dentro del análisis de legibilidad a diferentes temas curriculares.

La investigación relacionada con la utilización de fórmulas matemáticas para el análisis de legibilidad sobre material educativo en diferentes temas ha sido ampliamente atendida, para el idioma español. Uno de estos trabajos es (BLANCO PÉREZ; GUTIÉRREZ COUTO, 2002) new questions are now coming to bear, such as the quality of the information provided or its ability to convey knowledge. The purpose of this study is that of gauging the degree of readability of this information for patients which is accessible on the Internet in Spanish. Methods: A total of 112 webpages dealing with seven (7 el cual resalta la importancia que tiene el contenido en internet sobre información de la salud para la población general, por lo que su objetivo es medir el grado de legibilidad de la información disponible a los usuarios. Para esto seleccionaron 112 documentos relacionados con diversas enfermedades sobre los cuales utilizan las estadísticas de texto que ofrece Word y la fórmula de Fernández Huerta para medir el grado de legibilidad; con este método encontraron que los documentos contaban con una legibilidad del nivel de lectura de un niño de 13 años, concluyendo que, aunque los documentos son buenos no son óptimos para los usuarios que buscan información, por lo mismo recomiendan la elaboración de documentos con información adaptada al nivel de legibilidad fundamental para los pacientes.

Un caso similar es el planteado en (BALLESTEROS-PEÑA; FERNÁNDEZ-AEDO, 2013) el cual realiza un estudio descriptivo, sobre la información correspondiente a los 30 medicamentos más entregados en España, en este caso los autores utilizaron el índice de Flesch-Szigriszt como herramienta de evaluación y la escala de INFLESHZ como escala de legibilidad, a partir de los resultados donde solo 10 de los 30 medicamentos presentan una legibilidad aceptable; los autores logran concluir que las estructuras gramaticales de los apartados de indicaciones, contraindicaciones y efectos adversos, son generalmente complejas de comprender para el ciudadano promedio.

En el análisis de legibilidad de contenido web la utilización de herramientas web es fundamental, por esta razón en (DORUK *et al.*, 2020) utilizan www.readable.io para llevar a cabo el análisis de 26 páginas web relacionadas con lesiones benignas de las cuerdas vocales, a partir del estudio los autores logran



concluir que la calidad, la legibilidad y la comprensión relacionadas a este tipo de tema en las primeras páginas web proporcionadas por Google son muy bajas para sus usuarios, por lo mismo recomiendan la revisión del contenido.

En el entorno escolar se resalta la falta de comprensión que llegan a presentar los estudiantes en diversos temas, por lo que es común realizar análisis de legibilidad sobre el material educativo, generando métodos para llevar este tipo de análisis. En este contexto la evaluación realizada por PISA en 2012 establece que los estudiantes indonesios presentan en promedio una habilidad en estos temas del 56.2% por lo que en (RUSNAYATI; HERVIANA; SARI, 2019) se presenta un estudio de legibilidad de los materiales de enseñanza sobre física basados en alfabetización científica, utilizando el procedimiento de Cloze diseñado por Wilson Taylor (ARTOLA GONZÁLEZ, 1991), el cual consiste en omitir sistemáticamente palabras de un texto y evaluar la capacidad del lector al momento de reemplazar las palabras faltantes. Las pruebas establece una legibilidad del material educativo del 67.5%, llevándolos a deducir que el material presenta un alto nivel de legibilidad.

Por otro lado, en el material educativo en temas relacionados con la salud que se encuentran publicados en sitios web, las organizaciones de salud recomiendan que esta información cuente con una legibilidad de lectura de sexto grado, por lo que en (CORREA *et al.*, 2020)but it is not clear whether individuals can understand the information available online. Most health organizations recommend that health educational materials (HEMs) con el objetivo de analizar la legibilidad de este tipo de material educativo, utilizaron herramientas de cálculo de legibilidad automática para el análisis de tres índices Flesch Reading Ease Score, Flesch, Kincaid Grade Level y Simple Measure of Gobbledygook los cuales son comparados en función del tipo de material con este método, los investigadores logran evaluar 405 materiales identificando que el 93.8%, presentan niveles de legibilidad mayores al recomendado.

A partir de la revisión de trabajos relacionados, es posible observar que el estudio de la legibilidad de los textos puede emplearse con diversos enfoques curriculares debido a la gran cantidad de información disponible en la web. Sin embargo, se puede deducir que la finalidad de estas investigaciones es la creación de metodologías que permitan a través de cálculos matemáticos y herramientas tecnológicas establecer los niveles de complejidad de un texto, su tipo de publicación y el grado de escolaridad que se espera posea su lector, pero sin plantearse la necesidad de realizar este tipo de evaluación como un mecanismo previo a la creación del contenido web, ni como una herramienta para el mejoramiento de la accesibilidad en sitios web para personas con discapacidad, enfocadas a la lectura o simplemente la adecuación del contenido a usuarios con diferente nivel educativo.



4 PLANTEAMIENTO DEL PROBLEMA

El Gobierno Colombiano bajo el Plan de Lectura y Escritura planteado desde el Ministerio de Educación resalta como proceso fundamental dentro de la formación de los colombianos el desarrollo de habilidades lectoras con el objetivo de generar “ciudadanos activos”, es decir, colombianos con la capacidad de ser críticos en aspectos políticos, económicos, sociales y culturales (CULTURA, 2018b); bajo este enfoque se establece que la capacidad de comprender un texto es fundamental dentro del desarrollo social y el aprendizaje autodidacta.

El Instituto Estadístico de la UNESCO (UIS) en su último informe publicado sobre la comprensión lectora de los niños y jóvenes, indica que 6 de cada 10 niños en el mundo no leen bien y en América Latina y el Caribe al menos el 36% de los jóvenes culminan sus estudios básicos sin desarrollar las capacidades mínimas para ser críticos o apropiarse de un texto (UNESCO, 2017) la cual requiere una educación primaria y secundaria que lleve a resultados de aprendizaje relevantes y efectivos. A través del desarrollo de una nueva metodología y base de datos, el UIS ha producido una panorámica global de la situación de aprendizaje que enfrentan los niños y adolescentes que están dentro y fuera de escuela. Los datos muestran una necesidad crítica de mejorar la calidad de la educación, ampliando al mismo tiempo el acceso para asegurar que nadie se quede atrás. En el documento también se discute la importancia del análisis comparativo y el concepto de niveles mínimos de competencia. Más de 617 millones de niños y adolescentes no están alcanzando los niveles mínimos de competencia (NMCs). En el caso colombiano, los resultados no han sido los esperados, esto es demostrable en los resultados de las pruebas internacionales PISA (Evaluación Internacional de Alumnos de la OCDE) del 2018 en las cuales los jóvenes colombianos obtuvieron un rendimiento menor a la media de los países que conforman la Organización para la Cooperación y el Desarrollo Económico (OCDE) y en lectura su resultado fue menor al registrado en las pruebas del 2015 (MARKUS, 2019) el cual evalúa hasta qué punto han adquirido los conocimientos y competencias fundamentales para su participación plena en la sociedad. La evaluación se centra en el dominio de la lectura, las matemáticas y las ciencias y en un ámbito innovador (en 2018, este ámbito innovador fue la competencia global).

En estudios para identificar los hábitos de lectura y escritura de los colombianos en zonas urbanas y rurales, teniendo presente la influencia de la tecnología y las herramientas de información como parte fundamental de la vida cotidiana de los colombianos, se identificó que el 70.4% de las personas mayores de 5 años leen con mayor frecuencia en un soporte digital y el 38.6% utilizan páginas web como material de lectura, solo por debajo de redes sociales y correos electrónicos (DANE, 2018)(CULTURA, 2018a). Contar con este tipo de información sobre los hábitos de lectura de los colombianos, permite identificar la



necesidad de que al momento de crear o entregar contenido digital cuente con características acordes al público al cual está enfocado.

Una de las herramientas para determinar lo anterior, es el análisis de legibilidad lingüística del contenido textual de un sitio web, el cual es la medición de la facilidad o complejidad de lectura de un texto a partir de aplicación de modelos matemáticos, los cuales permiten identificar la relación existente entre las condiciones léxicas y gramaticales de un texto y la comprensión lectora (ROJAS VILLARROEL, 2019). Esto es un elemento imprescindible en la redacción del contenido textual de un sitio web que sea de carácter perlocutivo o educativo, ya que identificar el nivel de legibilidad lingüística puede llegar a ser considerado como un referente cuantitativo de la calidad de la información utilizada en el sitio web (CASTILLO, 2014).

Un aspecto relevante cuando se analiza el contenido disponible a través de la web, es determinar su nivel de accesibilidad. Para esto es posible realizar procesos de evaluación de accesibilidad web analizando el cumplimiento de las Pautas de Accesibilidad al Contenido Web (WCAG) (W3C, 2010). Sin embargo, existen diferentes dificultades al momento de realizar estos procesos de evaluación. Específicamente, cuando se trata de determinar el cumplimiento de los criterios contemplados dentro de la Pauta 3.1 que corresponde a la Legibilidad de los textos, los desarrolladores web no cuentan con mecanismos que permitan determinar que tan avanzado es el nivel de lectura requerido para comprenderlo y así poder agregar otras versiones del texto o contenido de apoyo. Los desarrolladores no logran determinar el nivel de legibilidad del contenido textual, según los niveles de educación definidos en la clasificación internacional de la educación (UNESCO, 2013).

Teniendo en cuenta lo anterior, se hace evidente la necesidad de proporcionar herramientas que apoyen el proceso de evaluación de accesibilidad durante la creación de sitios web, que permitan establecer el nivel de legibilidad lingüística del contenido textual, con el objetivo de determinar y esclarecer el público al cual está dirigida la página web, lo que permitiría tomar las acciones necesarias para dar cumplimiento a los criterios de accesibilidad y lograr sitios web más accesibles.

5 PROPUESTA

A partir del planteamiento presentado en la sección anterior, se propone una herramienta de análisis de legibilidad lingüística en español, que puede servir de apoyo en procesos de evaluación de accesibilidad a un sitio web. La herramienta propuesta recibió como nombre *Honkma* y se encuentra disponible para ser usada por el público en general, que desee analizar el nivel de legibilidad de un texto en español, utilizando diferentes fórmulas. En la Figura 1 se presenta la interfaz principal de la herramienta.



Figura 1 –Interfaz Gráfica de la herramienta Hohkma

Hohkma

Herramienta de apoyo para los procesos de evaluación de legibilidad lingüística.

¿Qué es Hohkma? ¿Qué es la legibilidad?

Contenido a Evaluar.

Introduce la dirección web que quieres analizar.

Evaluar

UNIVERSIDAD NACIONAL DE COLOMBIA

GRUPO DE AMBIENTES INTELIGENTES ADAPTATIVOS

Fuente: Elaboración propia

Hohkma usa un algoritmo que aplica los índices de legibilidad de Fernández Huerta y el índice de perspicuidad de Flesch-Szigriszt. Esta herramienta analiza la información siguiendo los pasos que se presentan a continuación:

1. Al pulsar el botón “Evaluar” comprueba si el texto introducido es una dirección o URI, si es correcto se descarga la página web que analizara y pasa al siguiente paso.
2. La herramienta solo funciona si se introduce un texto plano o un fichero marcado como XML o HTML. Si el archivo del sitio web no se encuentra en ninguno de estos formatos la herramienta generara un error.
3. El algoritmo elimina todas las etiquetas de XML y HTML (si las tiene), ya que estas no se toman en cuenta dentro del análisis de legibilidad y solo se toma el texto. Al tratarse de un sitio web, se deben eliminar aquella información no relevante para el análisis como: publicidad, menús, encabezados, pie de página, avisos de cookies, etc. Para llevar a cabo este proceso de utiliza Python Readability, librería de Python utilizada principalmente en el análisis de textos.
4. La herramienta cuenta la cantidad de caracteres dentro del sitio web ya que si cuenta menos de 100 caracteres el algoritmo de análisis no se ejecutará debido a que por definición los índices utilizados en esta herramienta no admiten un texto de menos de 100 caracteres.
5. La herramienta está pensada para el idioma español, para asegurar que solo se ejecuten resultados para este idioma se tomó la decisión de usar la biblioteca de detección de idiomas de Google a *Python Langdetect*.



6. Para generar los resultados de estadísticas del texto durante el proceso de evaluación, se cuentan los siguientes datos del texto: sílabas, párrafos y letras.
7. Una vez que el sitio web a cumplido con todos los requerimientos para ser evaluado y se prepararon los insumos necesarios para ejecutar los índices de legibilidad, se inicia el proceso de evaluación calculando los índices.
8. Por último se presentan los resultados, indicando el valor, nivel de dificultad y nivel educativo para cada índice, seguido de recomendaciones al respecto.

El tiempo de ejecución del proceso de evaluación de legibilidad esta dado por la cantidad de información a analizar, sin embargo, el tiempo máximo de respuesta es de 1 minuto si el texto analizado supera el millón de caracteres. En la Figura 2 se presenta la interfaz del módulo de evaluación que permite agregar el texto a analizar e iniciar el proceso y en la Figura 3 la visualización de los resultados de evaluación.

Figura 2 – Interfaz módulo de evaluación

Contenido a Evaluar:

Introduce la dirección web que quieras analizar.

Evaluar

Fuente: Elaboración propia

Figura 3 –Visualización de resultados de evaluación

Legibilidad del Texto			
Índice	Valor	Dificultad	Nivel Educativo
Fernández-Huerta	65	Normal	7° a 8°
Flesch-Szigriszt	55	Normal	6° a 9°

Fuente: Elaboración propia

La visualización de los primeros resultados dentro de la herramienta es a través de una tabla ubicada en la parte superior izquierda del cuerpo de la interfaz, esta tabla presenta el nombre del índice calculado, el valor calculado, el grado de dificultad y el nivel educativo. Esta información es obtenida a partir de los índices de legibilidad de José Fernández Huerta y Flesch-Szigriszt. Sin embargo, este último solo debido a la falta de consistencia en los grados de dificultad encontrada por Inés Cantalejo en (BARRIO CANTALEJO, 2007), se utiliza la escala de Inflesz para determinar el grado de dificultad y el nivel educativo del texto evaluado.



A partir, de la revisión y exploración de herramientas tanto online como offline para análisis de legibilidad, se encontró que el proporcionar estadísticas del texto evaluado es parte de los resultados comunes que proporcionan, siendo en unas más detalladas que otras; sin embargo, se observó que estas estadísticas no proporcionaban o representaban un componente adicional para el evaluador y en muchos casos debido a la gran cantidad de estadísticas ocasionaba que la herramienta fuera visualmente difícil de interpretar.

Hohkma proporciona de igual forma una serie de estadísticas sobre el texto evaluado, las cuales tiene como propósito apoyar al evaluar si este decide realizar una evaluación manual, ya sea de los índices de legibilidad entregados en la herramienta o de algún otro índice que requiera dicho dato. Esta funcionalidad presenta 3 estadísticas del texto: número de sílabas, número de palabras y número de frases.

En *Hohkma* se realiza explicación de los resultados, entregando información adicional sobre el resultado obtenido a partir del índice, esto con dos propósitos; el primero proporcionar una evaluación más completa del texto y segundo apoyar al evaluador si este no cuenta con conocimientos amplios en cuanto a la evaluación de la legibilidad lingüística. Las explicaciones generadas por la herramienta están clasificadas de acuerdo al nivel de dificultad generado por el índice.

Finalmente, *Hohkma* cuenta con una sección de recomendaciones, con el objetivo de mejorar el texto, reducir o aumentar la dificultad de legibilidad del texto a través de una serie de recomendaciones principalmente de forma. Las recomendaciones generadas por la herramienta están clasificadas de acuerdo al nivel de dificultad generado por el índice.

6 CONCLUSIONES

El análisis de legibilidad lingüística desde un enfoque accesible y utilizado en los procesos de calidad durante la construcción del contenido textual de un sitio web, aun es un tema poco tratado a nivel del idioma español y más aun en Latino América, donde es evidente la necesidad de herramientas que apoyen a los usuarios en el mejoramiento de la comprensión lectora.

El análisis de legibilidad desde un enfoque para el mejoramiento de la accesibilidad de los sitios web, aun no es muy empleado, pero puede llegar a ser una herramienta de gran apoyo en procesos de evaluación de accesibilidad. Además, las herramientas de análisis de legibilidad lingüística para el español, aunque cumple su propósito principal el cual es realizar una evaluación, cálculo de los índices de legibilidad y estadísticas gramaticales de los textos evaluados, deberían contar con un nuevo enfoque, donde se presente más como una herramienta de apoyo para el evaluador, al proporcionarle información concisa y



relevante para emitir un criterio sobre el nivel educativo necesario para comprender un texto, sin que este requiera un alto conocimiento sobre la legibilidad lingüística.

Se plantea como trabajo futuro, bajo el enfoque de la herramienta *Hohkma*, la necesidad de establecer recomendaciones más detalladas para el mejoramiento de contenido y contar con la opción de analizar textos en otros formatos. Además, se espera que la herramienta pueda tener un módulo de textos alternativas, es decir, generar de forma automática textos con mayor o menor nivel de legibilidad lingüística según lo solicite el usuario de la herramienta.

REFERENCIAS

ALLIENDE, Felipe. Evaluación de la legibilidad de los materiales escritos. **Lectura y Vida**. Newark, Delaware, IRA, [s. l.], 1990.

ANTONELLI, Humberto Lidio; FORTES, Renata Pontin de Mattos. A Support for Developers Implement the Accessibility Guidelines Regarding to Web Menus. **Proceedings of the 21st Brazilian Symposium on Multimedia and the Web - WebMedia '15**, [s. l.], p. 33–40, 2015. Disponible: <https://doi.org/10.1145/2820426.2820445>. Acceso en:

ARAGALL, Francesc. **La accesibilidad en los centros educativos**. 1. ed. Madrid: Cinca, S. A, 2010. v. 1.

ARTOLA GONZÁLEZ, Teresa. El procedimiento cloze: una revisión general. **Revista complutense de educación**, [s. l.], v. 2, n. 1, p. 69–82, 1991. Disponible: <https://doi.org/10.5209/RCED.19025>. Acceso en:

BAGUMA, Rehema; LUBEGA, Jude T. A web design framework for improved accessibility for people with disabilities (WDFAD). In: , 2008, New York, New York, USA. **Proceedings of the 2008 international cross-disciplinary workshop on Web accessibility (W4A) - W4A '08**. New York, New York, USA: ACM Press, 2008. p. 134. Disponible: <https://doi.org/10.1145/1368044.1368077>. Acceso en:

BALLESTEROS-PEÑA, Sendoa; FERNÁNDEZ-AEDO, I. Análisis de la legibilidad lingüística de los prospectos de los medicamentos mediante el índice de flesch-szigriszt y la escala inflesz. **Anales del Sistema Sanitario de Navarra**, [s. l.], v. 36, n. 3, p. 397–406, 2013. Disponible: <https://doi.org/10.4321/s1137-66272013000300005>.

BARRIO-CANTALEJO, Inés María *et al.* Validación de la escala INFLESZ para evaluar la legibilidad de los textos dirigidos a pacientes. **Anales del Sistema Sanitario de Navarra**, [s. l.], v. 31, n. 2, p. 135–152, 2008. Disponible: <https://doi.org/10.4321/s1137-66272008000300004>. Acceso en:

BARRIO CANTALEJO, Inés María. **Legibilidad y salud**. [s. l.], p. 70–96, 2007.



BLANCO PÉREZ, Azucena; GUTIÉRREZ COUTO, Uxía. Legibilidad de las páginas web sobre salud dirigidas a pacientes y lectores de la población general. **Revista Española de Salud Pública**, [s. l.], v. 76, n. 4, p. 321–331, 2002. Disponible: <https://doi.org/10.1590/s1135-57272002000400007>. Acceso en:

CASTILLO, Alonso monfilio guamán. La legibilidad de los textos escolares como elemento motivador del aprendizaje de los alumnos de educación general básica superior. [s. l.], p. 165, 2014.

CORREA, Daniel José *et al.* Quantitative readability analysis of websites providing information on traumatic brain injury and epilepsy: A need for clear communication. **Epilepsia**, [s. l.], v. 61, n. 3, p. 528–538, 2020. Disponible: <https://doi.org/10.1111/epi.16446>. Acceso en:

CULTURA, Ministerio de. **Las Bibliotecas, una apuesta para abrir caminos de paz en los lugares más apartados del país**. [S. l.], 2018a.

CULTURA, Ministerio de. **PLAN NACIONAL DE LECTURA Y ESCRITURA LEER ES MI CUENTO Nombre del Plan**. Bogotá: [s. n.], 2018b.

DANE. **Boletín técnico Encuesta Nacional de Lectura (ENLEC)**. Bogotá: [s. n.], 2018.

DORUK, Can *et al.* Readability, Understandability, and Quality of Online Education Materials for Vocal Fold Nodules. **Journal of Voice**, [s. l.], v. 34, n. 2, p. 302.e15–302.e20, 2020. Disponible: <https://doi.org/10.1016/j.jvoice.2018.08.015>. Acceso en:

FERNÁNDEZ, Alba; MANRIQUE-ABRIL, Fred; BAUTISTA SAAVEDRA, Camilo. **ACEPTACIÓN POR EXPERTOS Y LEGIBILIDAD DE MATERIAL ESCRITO Y AUDIOVISUAL: CALIDAD Y PROPIEDADES PSICOMÉTRICAS21**. Pereira: [s. n.], 2010. Disponible: <https://doi.org/0124-8146>. Acceso en:

FIGUEROA, Marcia V. Sappa *et al.* **Evaluación de la Accesibilidad en Dos Sitios Bancarios Nacionales Dependientes de la Administración Pública**. [s. l.], v. 2, n. 3, p. 144–148, 2014.

GARCÍA, Miguel Ángel; ARÉVALO, César; HERNÁNDEZ SUÁREZ, Mayra. La comprensión lectora y el rendimiento escolar. **Cuadernos de Lingüística Hispánica**, [s. l.], v. 32, n. 0121-053X, p. 155–174, 2018.

GUAMÁN CASTILLO, Alonso Monfilio. **La Legibilidad de los Textos Escolares como Elemento Motivador del Aprendizaje de los Alumnos de Educación General Básica Superior**. 1–164 f. 2014. - UNIVERSIDAD NACIONAL DE PIURA, Piura, 2014.

HERNÁNDEZ, Iván Ríos. Influencia del lenguaje y lugar de origen de un lector en la comprensión de mensajes de comunicación en salud y en la formación de conductas preventivas. **Observatorio (OBS*)**, [s. l.], v. 4, n. 1, p. 185–200, 2010. Disponible: <https://doi.org/10.15847/obsOBS412010309>. Acceso en:



HILERA, José R. *et al.* Evaluación de la accesibilidad de páginas web de universidades españolas y extranjeras incluidas en rankings universitarios internacionales ; Accessibility assessment of web pages of Spanish and foreign universities included in international rankings. **Revista Española de Documentación Científica**, [s. l.], v. 36, n. 1, p. 1–16, 2013.

KIRKPATRICK ANDREW, Cooper Michael. **WCAG 2.1 es una recomendación del W3C**. [S. l.], 2018.

MARKUS, Schwabe. **Programme for International student assessment (PISA) Results From PISA 2018**. [S. l.: s. n.], 2019.

MIÑÓN, Raúl *et al.* An approach to the integration of accessibility requirements into a user interface development method. **Science of Computer Programming**, [s. l.], v. 86, p. 58–73, 2014. Disponible: <https://doi.org/10.1016/j.scico.2013.04.005>. Acceso en:

MUÑOZ BAQUEDANO, Miguel. Legibilidad y variabilidad de los textos. **Corporación Municipal Viña del Mar para el Desarrollo Social**, [s. l.], v. 813, p. 1–13, 2006.

PORRAS-GARZÓN, Jorge M; ESTOPÀ, Rosa. Círculo de Lingüística Aplicada a la Comunicación. [s. l.], p. 205–216, 2020.

RIOS HERNANDEZ, Ivan Neftali. Un acercamiento a la legibilidad de textos relacionados con el campo de la salud. **Chasqui. Revista Latinoamericana de Comunicación**, Quito, Ecuador : CIESPAL, n. 135, p. 253, 2017. Disponible: <https://doi.org/10.16921/chasqui.v0i135.2892>. Acceso en:

ROJAS VILLARROEL, Daniel. **Lectorabilidad y géneros del conocimiento: análisis del texto escolar de lenguaje y comunicación de octavo año básico en Chile**. [s. l.], p. 2–51, 2019.

RUSNAYATI, H.; HERVIANA, G.; SARI, I. M. Analysis of scientific literacy-based junior high school physics teaching materials readability on students increasing scientific literacy skills. *In:* , 2019. **Journal of Physics: Conference Series**. [S. l.]: Institute of Physics Publishing, 2019. Disponible: <https://doi.org/10.1088/1742-6596/1280/5/052010>. Acceso en:

SALAZAR GRIJALVA, Daniel. **Análisis de la accesibilidad y usabilidad gráfica en los sitios web gubernamentales de la provincia de Esmeraldas para personas con discapacidad visual y auditiva**. 63 f. 2019. - PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR SEDE ESMERALDAS UNIDAD, [s. l.], 2019.

UNESCO, Instituto de estadística de la. **Clasificación Internacional Normalizada de la Educación (CINE)**. [S. l.: s. n.], 2013. Disponible: <https://doi.org/10.15220/978-92-9189-129-0-spa>. Acceso en:



UNESCO. Más de la mitad de los niños y adolescentes en el mundo no ésta aprendiendo. **Uis Instituto De Estadística De La Unesco**, [s. l.], v. 0, n. 46, p. 18–26, 2017.

W3C. **Comprender el Criterio de Conformidad 3.1.5 | Comprender las WCAG 2.0**. [S. l.], 2008.

W3C. **Comprender la Pauta 3.1 | Comprender las WCAG 2.0**. [S. l.], 2010.